

## 6. DATA MANAGEMENT AND ARCHIVING

The data management mission of the GRYN is to provide data and information resources that are organized, available, useful, compliant and safe. To achieve these fundamental requirements, the data management plan focuses on the following objectives:

- Provide data management services and guidance in support of the I&M program goal to identify, catalog, organize, structure, archive and make available relevant natural resource information
- Initiate and invest in data management activities based on data and information needs defined in network monitoring protocols and inventory study plans
- Integrate data management activities with all aspects and at all stages of network business
- Specify data stewardship responsibilities for all personnel
- Collaborate internally and externally to address data management issues with individuals at all organizational levels.

The I&M program provides a framework for natural resource information management (Figure 6.1) aimed at achieving maximum returns on investments made in data gathering, such that relevant

data and information is available long term to multiple levels of the organization (park, network, regional, national). This framework includes these elements:

- Provide standards for natural resource inventories
- Develop and support Service-wide online natural resource database applications
- Provide desktop database applications that mirror master databases and promote standard data entry and organization
- Recommend a natural resource database template that allows local flexibility but also promotes design consistency for the purpose of sharing database designs and content
- Direct networks to hire data management staff and emphasize data management
- Mandate written network data management plans
- Require written data management procedures and responsibilities in inventory study plans and vital sign monitoring protocols.

The network builds on this framework by applying data management guidelines for a monitoring vision (Figure 6.2) that is directly related to

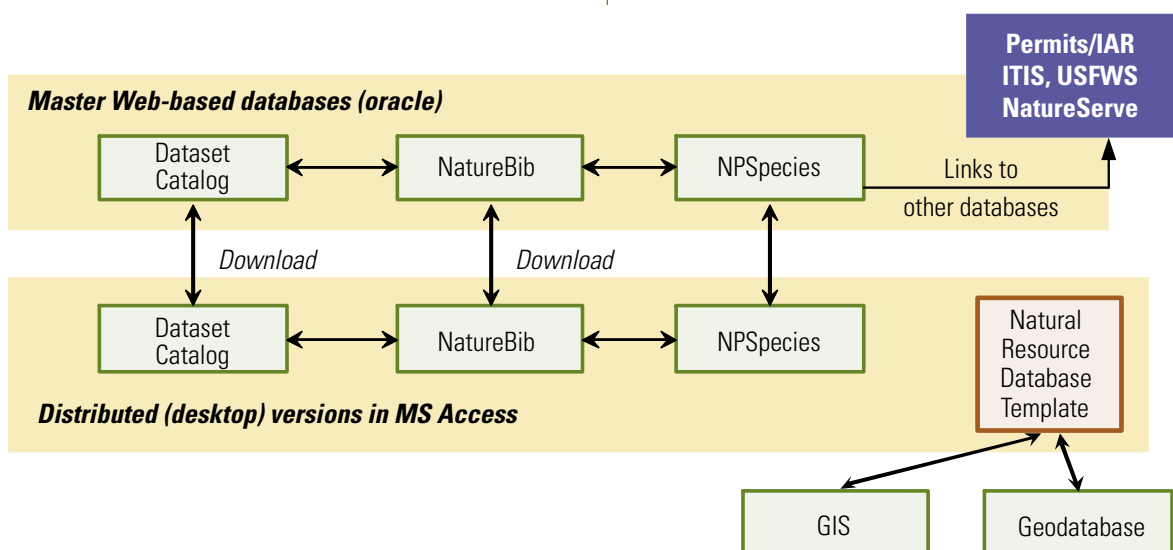


FIGURE 6.1 Integrated natural resource data management framework.

the needs of local scientists and natural resource managers. The network manages all forms of inventory and monitoring data and information, provided they support one or more of the following program goals:

- Goals and objectives of the Inventory and Monitoring program
- Specific information needs defined in approved Vital Sign monitoring protocols
- Network inventory study plan objectives
- Other specific natural resource management projects that network personnel and park staff agree to cooperate in developing and managing.



**FIGURE 6.2** Monitoring vision (adapted from National Water Quality Monitoring Council 2004).

The Network Data and Information Management Plan (Appendix VIII) outlines the strategy and guidelines for thorough, integrated and coordinated resource information management activities that attempt to link Service-wide information requirements and data management tools with park-level inventory and monitoring information needs.

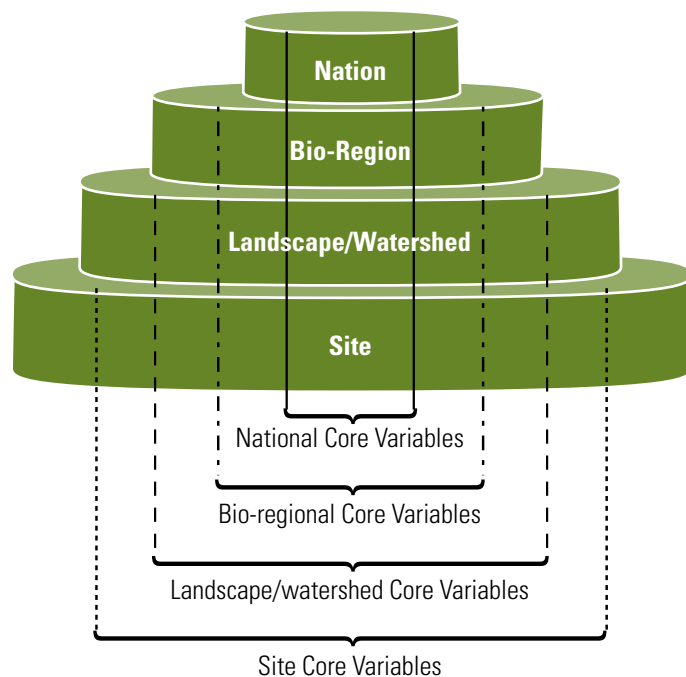
### *Roles and Responsibilities for Data and Information Stewardship*

The benefits to managers and scientists from inventory and monitoring projects are substantially affected by the ability to track data from the time they are gathered until and while they inform a decision-making process. In many cases this involves time frames of several years and includes changes in information technology, turnover in staff, new scientific insights and shifting priorities. The purpose of data stewardship is to share the responsibility for managing data and information resources. The network works to ensure mutual accountability for specific tasks (responsibilities) assigned to each position (role) involved with data as a producer, analyst, manager or end user. Table 6.1 lists primary roles and responsibilities for all data steward roles, some of which are commonly assigned to a single position, e.g. a resource specialist may serve as project leader. Individual monitoring protocols and inventory study plans draw on a more complete Data Stewardship Framework (Appendix VIII) for guidance in assigning specific jobs and detailed tasks related to data management.

**TABLE 6.1** Programmatic roles and responsibilities for data stewardship

Role	Programmatic Data Stewardship Responsibilities
Project Crew Member	Collect, record and verify data
Project Crew Leader	Supervise crew and organize data
Data/GIS Specialist or Technician	Process and manage data
Information Technology Specialist	Provide IT/IS support
Project Leader	Oversee and direct project operations, including data management
Resource Specialist	Validate and make decisions about data
GIS Manager	Support park management objectives with GIS and resource information management
Network Data Manager	Ensure inventory and monitoring data are organized, useful, compliant, safe and available
Database Manager	Know and use database software and database applications
Curator	Oversee all aspects of the acquisition, documentation, preservation and use of park collections
Statistician or Biometrician	Analyze data and present information
Network Ecologist	Integrate science in network activities
Network Coordinator	Coordinate and oversee all network activities
I&M Data Manager (National Level)	Provide Service-wide database availability and support
End Users (managers, scientists, publics)	Inform the scope and direction of science information needs and activities. Apply data and information services and products

The network data manager plays a fundamental role as coordinator of data management roles and activities. This involves understanding program and project requirements, developing and maintaining data management infrastructure and standards, and communicating with all responsible individuals. Integration and communication with GIS staff, natural resource information managers and I&M project leaders promotes common understanding and efficiency. Integration may include such activities as training, guidance and assistance for inventory and monitoring efforts and, where practical, for park stewardship requirements related to the broader realm of Service-wide natural resource information management. The network data manager works closely with each project leader to meet the data management requirements specified in monitoring protocols and inventory study plans. This includes substantial involvement in project planning, crew training, field work, progress and deliverable tracking, and other relevant project operations.



**FIGURE 6.3** The “wedding cake” model of variables (Powell 2000).

### *Data Management Program Overview*

Data and information management in the GRYN will attempt to support an adaptive yet consistent approach to managing and delivering a useful suite of natural resource inventory and monitoring data

and information. The network relies primarily on the general and interrelated data and geodatabase models of Microsoft® Access and Environmental Systems Research Institute (ESRI®) GIS software applications. Department of Interior (DOI) Enterprise Resource Management efforts and the NPS GIS and natural resource data management communities are heavily invested in these products and tools like GIS Theme Manager, AlaskaPak and the NPS Metadata Tools and Editor. Since Access databases and ESRI personal geodatabases can be scaled to enterprise level solutions as the Service continues to develop its corporate information management strategy, it is appropriate for the network to use these data models to meet objectives for managing and delivering monitoring data. The network expects Service-wide information needs (the core of the wedding cake in Figure 6.3) to be identified, coordinated and addressed at the highest levels of NPS. As these institutional requirements and solutions for enterprise business needs evolve, the network will continue to coordinate with internal and external stakeholders on data management activities.

### *Information Technology (IT) Infrastructure for Data Management*

The organization, availability and security of data and information resources depend on a solid computer system foundation. Where possible, the Greater Yellowstone Network uses DOI and NPS solutions for computer network, hardware and software requirements for data and information management, including: DOI Enterprise Resource Management hardware licensing for desktop, laptop and server hardware; virtual private network access; client and server operating systems; asset management software; email; security (Antivirus); desktop office and publishing; image processing; database and geographic information system applications.

The network is self-supporting for many of its IT needs. Most aspects of the network’s local computer system are managed by network staff using assistance and services from regional and national IT personnel to support and maintain an updated system. Park computer support staff also provide assistance when requested by the network.

Network data and information resources are compiled, organized and archived using a structured file system on a local server that is backed up to tape and Network Attached Storage (NAS) devices (Figure 6.4). The network staff will incorporate and follow NPS information technology policies, standards, procedures and guidelines available from the Office of the Chief Information Officer.

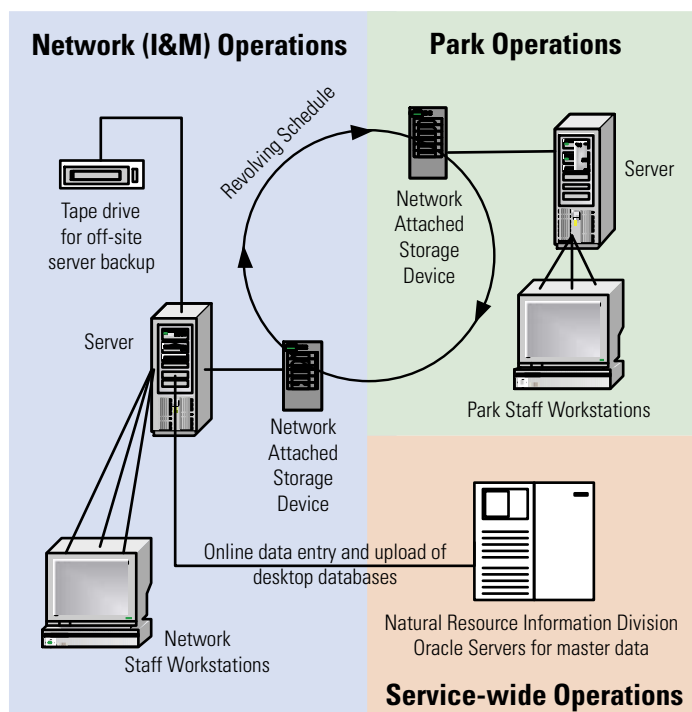


FIGURE 6.4 IT Connectivity Diagram

### Acquiring and Processing Data

The network's three-phase planning process resulting in peer-reviewed vital sign protocols provides a robust link between specific measurable information needs and user requirements. A critical result of this process is that data required for the long term monitoring of park vital signs are distinguished from all other existing and potentially new data sources. This provides a manageable scope for acquiring, processing and administering monitoring data. In order to provide a synthesis of scientific information based on vital signs and related data, the network also gathers and processes relevant data and information from other park-based and external inventory and monitoring efforts.

Past investments in data gathering in the GRYN have resulted in a legacy of products that vary widely in format, consistency and value for park stewardship (GRYN 2000). Future work and expense to link legacy data with management requirements must be carefully scrutinized by a group of professionals representing management, science and technical branches of park stewardship. Although GRYN-funded inventories have been completed, parks will continue to perform inventories according to the spirit and goals of the Natural Resource Challenge. The network expects to coordinate with these projects to preserve existing partnerships and integrate data management activities. To help address the volume of natural resource data stored at the parks, the network directly supports annual work by park staff

to obtain, catalog, report and archive data in NPSpecies, NatureBib and Dataset Catalog. This involves working with research permit and reporting staff, park natural resource specialists and external researchers to receive and compile new data as well as discover and process existing data sources.

New data are only acquired and processed by the network if they support specific objectives outlined in one of the following plans:

- Greater Yellowstone Network Vertebrate and Vascular Plant Inventory Study Plan (GRYN 2000)
- Monitoring protocols for vital signs listed in chapter three of this monitoring plan
- Reporting requirements listed in chapter seven of this monitoring plan
- Service-wide natural resource inventories
- Relevant projects with management-approved work plans in which the network is one stakeholder and contributor.

TABLE 6.2 Abbreviated Data Development Model

1. Identify issues and concerns
2. Define the purpose and need for data collection and analysis
3. Develop explicit monitoring objectives or inventory criteria (these are key questions addressing the issue or concern within the scope of the purpose and need)
4. List measurable, observable and predictable variables associated with each key question
5. Formalize and document information needs
6. Develop a data dictionary for field names, lists of values, quality factors and metadata characteristics
7. Select or develop an appropriate sample design
8. Identify and assign explicit data stewardship roles and responsibilities
9. Write a complete monitoring protocol or inventory study plan
10. Design or adopt/adapt a database (including quality control elements)
11. Plan for data acquisition (**beginning of data life cycle**)
12. Collect data— field and office components
13. Process data (includes verification, transfer, addition of required attributes)
14. Store, organize and secure data
15. Use, analyze and report data
16. Maintain and serve data and derived products
17. Archive data (long term storage that may require media and/or platform transfer)
18. Delete data that are no longer needed, if appropriate (**end of data life cycle**)

### Ensuring Data Quality

The network approaches quality assurance as “an integrated system of management activities involving planning, implementation, documentation, assessment, reporting and quality improvement to ensure that a process, item or service is of the type and quality needed and expected by the consumer” (Palmer 2003). The network strives to achieve appropriate data quality by:

- documenting requirements for data quality
- implementing data quality assurance activities in all stages of network operations
- using relevant quality control procedures throughout the data life cycle
- incorporating, teaching, and applying direction from NPS Director’s Order #11B: “Ensuring Quality of Information Disseminated by the National Park Service.”

The specific observed, measured or predicted elements that address the objectives for each vital sign monitoring protocol or inventory study plan include documented quality factors in the data dictionary for that plan. Once these are defined and recorded (step 6 from Table 6.2) appropriate quality assurance procedures can be applied. These include:

- training and awareness in quality assurance
- equipment selection, calibration and maintenance
- data collection procedures and data entry controls
- automated and user-assisted data verification routines
- user-assisted data validation routines
- pertinent quality controls based on water quality data collection and processing procedures.

### Documenting Data

Documenting data sets, the data source(s) and the methodology by which the data were acquired establishes the basis for interpreting and appropriately using data. The network requires the following documentation elements as insurance to protect investments in data gathering.

1. Feature-level metadata— characteristics about each feature/record in a database. Data records collected according to network protocols will include the name, date and version of the associated protocol. This is an example of feature-level metadata that promotes the longevity and utility of a data asset.
2. Data set metadata— documentation meeting Federal Geographic Data Committee and NPS standards.
3. Notes from field, laboratory and analysis work.
4. NPS Dataset Catalog records for brief metadata on all data hold-

ings, including locally published geospatial data sets (themes and images).

5. Monitoring plan or inventory study plan with complete background, objectives and methods that directly relate to the metadata and vice versa.

To achieve the required content and detail for metadata, the network uses a number of techniques, including:

- FGDC metadata content standards for geospatial and biological data sets
- specific metadata requirements outlined in task agreements and contracts with cooperators
- NPS training, online resources and software tools including Dataset Catalog
- educating personnel about roles and responsibilities for data documentation
- tracking project, data set and metadata status
- following up with responsible individuals to complete metadata
- incorporating feature level metadata into data gathering procedures.

The network creates, maintains, and publishes metadata according to NPS metadata standards and guidelines. All network metadata records meeting FGDC content standards for digital geospatial metadata are available on the NPS NR-GIS Metadata Clearinghouse Web site.

### Summarizing and Analyzing Data

Providing meaningful results from data summary and analysis is a cornerstone of the I&M program and characterizes the network’s data management mission to provide useful information for managers and scientists. Each monitoring protocol establishes requirements for on-demand and scheduled data analysis and reporting. Based on these requirements, the associated database(s) for the protocols include functions to summarize and report directly from the database as well as output formats for import to other analysis software programs. In addition to tabular and charted summaries, the network provides maps of natural resource data and GIS analysis products to communicate spatial locations, relationships and geospatial model results. Please see Chapter 7 for a more detailed description of the network’s analysis and reporting schedule and procedures.

### Distributing Data and Communicating Information

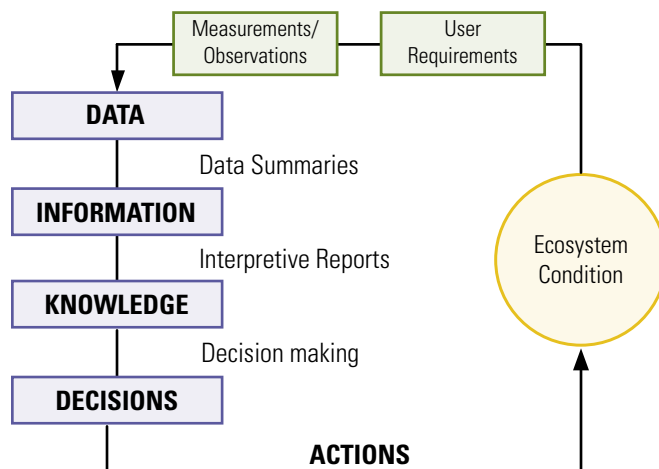
The network uses a variety of means to obtain, secure and share all network-generated data and scientific information while protecting the integrity and privacy of sensitive or protected data. The network’s Web site provides an information portal that assembles and links existing

and planned internet services that provide for most of the network's data and information distribution requirements. In the following list, access to virtually all network data and information represented in Figure 6.5 is permitted according to the security level of the user. All data are available to network parks, most data are available Service-wide, and non-protected data are available to all external users.

- Inventory and monitoring planning and project reports are online at the GRYN Web site.
- Park and network monitoring protocols and database designs are online at the NPS Protocol Clearinghouse.
- Searchable metadata are online at the NPS Natural Resource and GIS Metadata Database.
- Original and processed data sets from the parks and network are online at the NPS Biodiversity Data Store and/or the NPS Natural Resource and GIS Data Store.
- Annually updated water quality data are online at the Environmental Protection Agency's STORET Web site.
- Biodiversity data and information are available online at the NPSpecies Web site.
- Scientific citations are online at the NatureBib Web site.

Data sharing between the network and parks is scheduled and coordinated to ensure data in useful formats are regularly available to park natural resource managers. Until the GRYN server and park servers can share data on the same computer network, data transfer occurs via electronic file transfer, exchange of digital media and a system of revolving external hard drives or NAS devices.

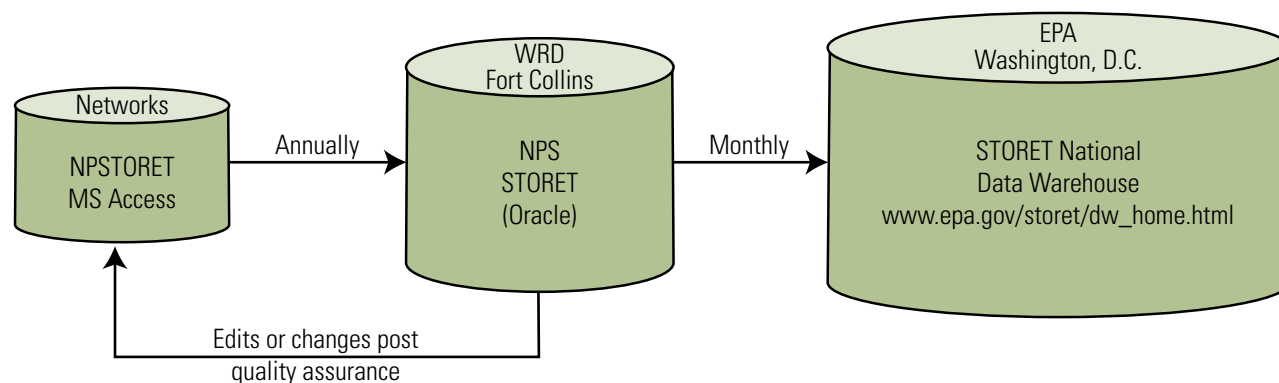
The network also serves data requests using file transfer protocol (FTP), attaching reports and other products with small file sizes to email and shipping digital media such as DVD, CD-ROM, diskette and magnetic tape cartridge.



**FIGURE 6.5** Decision Support Model (adapted from Palmer 2003).

### *Maintaining, Storing and Archiving Data and Information*

The data distribution mechanisms discussed above also provide data storage and archiving solutions. In addition to posting and submitting data to NPS and external data stores, the network maintains an organized file system stored on local server hard disks as well as backup media that includes optical, tape and Network Attached Storage devices. As future evolution of storage technology permits or demands, the full complement of network data will be migrated to new platforms and storage media. The maintenance requirements and associated roles and responsibilities for a given data set are specified in the monitoring protocol or inventory study plan and in the associated metadata. The network data manager keeps track of data maintenance schedules and works with project leaders and GIS specialists to review data maintenance requirements. Where nec-



**FIGURE 6.6** Natural Resource Challenge vital signs water quality data flow.

essary, the network acquires and archives data sets from multiple sources, but also relies on external repositories as the master data store for some ancillary data.

### *Water Quality Data*

Water quality data, including macroinvertebrate characteristics, are managed according to guidelines from the NPS Water Resources Division (Figure 6.6). This includes using the NPSTORET desktop database application at the parks to help manage data entry, documentation and transfer. The network oversees the use of NPSTORET according to the network's integrated and regulatory water quality monitoring protocols and ensures the content is transferred at least annually to NPS Water Resource Division for upload to the STORET database.



